



TITLE:

# 蛋白質折り畳みの計算機シミュレーション(数学者のための分子生物学入門,研究会報告)

AUTHOR(S):

岡本, 祐幸; 近藤, 剛史

---

CITATION:

岡本, 祐幸 ...[et al]. 蛋白質折り畳みの計算機シミュレーション(数学者のための分子生物学入門,研究会報告). 物性研究 2003, 81(1): 93-119

ISSUE DATE:

2003-10-20

URL:

<http://hdl.handle.net/2433/97611>

RIGHT:

蛋白質折り畳みの計算機シミュレーション  
岡崎国立共同研究機構・分子科学研究所  
総合研究大学院大学・数物科学研究科（併任）  
岡本 祐幸（おかもと ゆうこう）

§ はじめに

ノート作成 近藤剛史（京都大学大学院理学研究科）

拡張アンサンブル法という手法について解説する。計算機シミュレーションという言葉の意味は、計算機を使って自然現象を模擬的に再現する研究手法のことである。物理とか自然科学の研究手法に大きく分けて実験と理論があるが、計算機実験というのは、理論にかなり近いが、実験の要素も持っている。

蛋白質はどこで合成されるかというところ、ここ（図1のリボソーム）で合成されている。細胞があって、核がある。核の中に染色体があって、そこに遺伝情報、DNAがある。そこにアミノ酸配列の情報が書き込まれていて、そのアミノ酸配列の情報がメッセンジャーRNAにおくられて、そのメッセンジャーRNAがリボソームに行って、アミノ酸配列の情報に翻訳されて、その情報にしたがって、蛋白質が出来上がる。

蛋白質は自然界で機能を発揮するときは一定の形を持っている。合成されるというのはその形になるということである。いろいろに折り畳まれていて、一定の形になって、機能を発揮するのである。

染色体は、人間の場合は23対ある。一本の染色体は実はDNAの二重らせんになっている。それでそこに塩基の配列がアミノ酸配列として保存されている。

1980年代の後半に、日本人が、「ロボットを使って、自動的に塩基配列を読んでもいい。機械でやれば、ものすごくたくさんある情報を自動的に読んでいける」ということを提唱した。それで最初日本でやろうとしていたのだが、国を挙げてやろうと言うところまでには行かなかった。ところが、その間に、このアイデアにアメリカが飛び付いて、アメリカが中心になって進められる形になってしまった。後から元祖の日本がこの国際共同研究に参加するという形になったのである。

もう少し詳しく説明すると、DNAの二重らせんがからまったものが染色体だが、DNAを詳しく見ると、糖の部分とリン酸の部分と塩基の部分があって、この3つの部品からできていて、これをヌクレオチドという。この塩基の部分は自然界に4種類、アデニンとグアニンとシトシンとチミン、RNAの場合はチミンの代わりにウラシルでちょっとだけ違う。このヌクレオチドが鎖のようにつながったものをDNAという。丁度、塩基の部分は相棒がいて、水素結合でつながる相棒が、AだとTだし、GだとC、だから片側の配列が決まると、自動的に二重らせんの相棒の配列が決まる。これがしっかりと水素結合で結び付き、それがさらにらせんになって遺伝情報が正しく保存されている。

模式的には（図2）のようだが、実際はらせん構造をしている。形が決まるというのは不思議なくらい複雑な分子だが、これは1950年代に、DNAの結晶をつくって、それにX線をぶつけて回折の像を解析することによってこの形が決まったので

ある。

人間の場合は全部で 30 億個の塩基対がある。以前は 10 万個位の遺伝子があると言われていたがヒトゲノム計画で分かったことは意外に少なく、全部で 3 万～4 万個の遺伝子しかない。

## § ポストゲノム時代の最重要課題

ヒトゲノム計画でアミノ酸配列の情報が読み終わる。すると次に何をしないといけないかという、ポストゲノム時代はヒトゲノム計画によって決定された DNA の塩基配列 (すなわち、蛋白質のアミノ酸配列) の情報から、それぞれの蛋白質がどういう立体構造 (形) をして、どういう機能 (働き) をするかを探ること。私がやっていることがそうなのだが、アミノ酸配列の情報から、立体構造を予測する。それを計算機シミュレーションでやろうとしている。

## § アミノ酸と蛋白質

アミノ酸の側鎖に 20 種類の可能性がある。その側鎖を R と書くと、例えば R が水素原子だけだとグリシンになる。

二個のアミノ酸から水がとれてつながったものをペプチド結合という。ペプチド結合は平面構造をしている。

四つの原子がつながっていたときに、最初の 3 つが定める平面と、次の 3 つがつくる平面の成す二面角が決まる。

蛋白質の立体構造は、蛋白質を構成しているすべての原子の 3 次元座標をとることもできるが、長さや回転の自由度の角度を指定することによって、立体構造を規定することもできる。よって、蛋白質の立体構造は二面角の関数と言ってもいい。

図 3 は 1950 年代の終り頃、X 線の実験で回折データから立体構造が世界で初めて得られた、ミオグロビンという蛋白質である。ミオグロビンは筋肉にあって、酸素を貯蔵している。ちなみに、良く知られているヘモグロビンはこのミオグロビンが 4 個合わさったような形をしている。

ペプチド結合は平面構造をしていたので、この平面の部分をなぞっていくと、蛋白質の立体構造の一部がきれいならせん構造をしている場合がある (図 4)。この蛋白質の主鎖のらせん構造をアルファヘリックスという。らせん構造以外にもう一つあるのが、ベータシートという構造である (図 5)。

現在実験で決められた蛋白質の立体構造が 2 万個くらいあるが、その部分構造を見ると、統計的に、約 4 割がアルファヘリックス、2 割くらいがベータシート、他はそれ以外の形をしている。

アルファヘリックスだけのもの、ベータシートだけのものもあるが、アルファヘリックスとベータシートが共存しているものもある。

大体普通のサイズの蛋白質は 200～300 個くらいのアミノ酸からなる。一番小さ

いのは 50 個くらいで、最大の蛋白質は 6000 個くらい。300 個くらいのシミュレーションをしたいが、今やっているのは 37 個。水を入れないといけないので、ほとんど水のシミュレーションになってしまう。

## § 研究目的

なぜ蛋白質の立体構造に興味があるかというと、機能に深く関係しているからである。蛋白質にはいろいろな生化学的機能がある。酵素反応、酸素の運搬、貯蔵等。私が目指しているのは、ゲノムの情報つまりアミノ酸配列の情報のみを使い、ランダムな初期構造から出発して、計算機シミュレーションを使って、自然の立体構造を予測することである。そうすると立体構造が知られていない蛋白質の立体構造を計算機シミュレーションで知ることができる。

基礎研究としての興味は、蛋白質の折り畳み原理、機構の解明 (Protein Folding Problem) や、蛋白質の機能発現原理の解明であり、応用への発展性としては、新薬品の設計、特定の機能を持った人工蛋白質の設計、蛋白質の誤った折り畳みに起因する病理の研究 (例えば、狂牛病やアルツハイマー病) などがある。プリオン (狂牛病の原因となる蛋白質) は、本来ベータシートではないはずの部分の構造が何らかの原因でベータシートの構造をとることによって発病することが分かっている。

## § Anfinsen の実験

蛋白質の折り畳みを研究するために、計算機シミュレーションで細胞 1 個を扱うのは不可能である。その中にはものすごくいろいろなものが入っている。もちろん模型を非常に簡略化すればできるが、一つ一つの原子を扱おうとすれば、蛋白質だけでも複雑である。計算機シミュレーションでは絶望的にみえる。ところが、幸いなことに、1960 年代の初頭に、Anfinsen が次のような実験をした。

いわゆる試験管内の実験で、細胞から取り出した蛋白質に、変性剤を加えて、固まっていたのをぐちゃぐちゃにほどく。それで変性剤を取り除くと、ぐちゃぐちゃになっていたのが元の形に戻ることを実験的に示した。すると、細胞内のもろもろの複雑な環境は忘れて、蛋白質一分子と、周りの水の環境さえ用意すれば良いことになる。これなら計算機シミュレーションでも行けるはずだと、案外いけるんじゃないかという希望を与えた。

これは 60 年代初頭だが、40 年経ってもまだなかなかうまく行かない。難しい問題というふうに言われている。

一般に蛋白質は加熱すると戻らないと思われるが、それは他の蛋白質と絡まってしまうからである。単体 (一分子) だと元に戻るのである。

つまり、ほどかれてぐちゃぐちゃになった状態の蛋白質 (ランダムコイル) にはアミノ酸の配列が何だったかという情報しか残っていない。形の情報はぐちゃぐちゃになってしまった。それが元に戻るということは、アミノ酸の配列の情報に、蛋白

質の立体構造の情報が、含まれているということになる。これを Anfinsen's Dogma という。

この事は、蛋白質の折り畳みを研究している、実験、理論、両方の人たちにとって、過去 40 年に渡って、指導原理になってきた。これで Anfinsen はノーベル賞をとっている。

このことから、系の正しいエネルギー関数が与えられれば、後は、計算機シミュレーションでエネルギー最小状態を求めれば、蛋白質分子の立体構造を予測できるであろうと考えられる。

## § エネルギー関数とシミュレーション手法

計算機で蛋白質分子の立体構造を予測するには、エネルギー関数がしっかりしていないとダメである。さらにシミュレーション手法もしっかりしていないとダメ。何故かという、蛋白質のように非常に大きな、多自由度の系だと、系にエネルギー極小状態が無数に存在する。計算機シミュレーションやると、すぐ、そのエネルギーローカルミニマムに引っ掛かってしまう。グローバルミニマムになかなか行かない。

大体蛋白質が折り畳まれるのに、ミリ秒から分の時間がかかると言われている。ミリ秒から分と言うと、これはすごく長い時間スケールである。分子振動の時間スケールは  $10^{-14}$  秒とか、それくらいの時間で分子振動している。それなのに蛋白質がランダムコイルから折り畳まれるときに、ミリ秒とか分とか、そういう時間が必要だということは、ものすごく長い時間である。それはそれだけエネルギー極小状態がたくさん存在するということである。ただ実験でミリ秒とかはすぐに終わってしまうが...

エネルギー関数とシミュレーション手法のふたつをしっかりしたものにしよう、ということを目指している。

## § エネルギー関数

厳密なエネルギー関数を用いる。

まずエネルギー関数だが、蛋白質自身の構造とその周りの水の存在、このふたつがある。それで、このエネルギーで、原子一個ずつに取り入れたシミュレーションに標準的に使われているのがこの形である。

蛋白質自身の標準的な構造エネルギーは、ボンド長項 + ボンド角項 + ねじれエネルギー項 + Lennard-Jones 項 + 静電相互作用項で表されるので、全体で以下のように書くことができる。

$$E_{total} = \sum_{bonds} K_B(\ell - \ell_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2} \{1 + \cos(n\phi - \gamma)\} \\ + \sum_{i < j} \left( \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right) \\ + \text{溶媒との相互作用エネルギー}$$

ここで、 $\ell$  は原子と原子を結んでいる共有結合のボンドの長さである。その値は、大体実験的に分かっている、それを今  $\ell_0$  と呼んでいる。それぞれのボンドの長さの所が、実験で分かっている値になるように、 $\ell_0$  の値をボンドごとに入れてやるわけである。つまりそれぞれのボンドの長さ  $\ell$  が  $\ell_0$  になったとき、0 になり、 $\ell$  が  $\ell_0$  から離れると、当然エネルギーは大きくなってしまふ。エネルギーが一番小さい、一番安定な構造は、 $\ell_0$  の長さに丁度なっているときなのである。

ボンド角というのは、ボンドとボンドの成す角度であり、これを  $\theta$  とおくと、これも実験的にほとんど固定されている。エネルギーを小さくしようとすると、角度はその実験的に知られている値  $\theta_0$  になるようになっている。これらのバネ定数は、中性子をおつけて中性子散乱をみて、その散乱データから求めることができる。アミノ酸の、それぞれの形をまず規定するようなボンドの長さ、ボンドの角度を規定する。

次に Lennard-Jones 項というのがある。  $1/R_{ij}^{12} - 1/R_{ij}^6$  という形をしている。ここで、 $R_{ij}$  は原子  $i$  と原子  $j$  の間の距離である。これは普通、分子間力、ファンデルワールス力と呼ばれているものに対応する。この  $1/R_{ij}^{12} - 1/R_{ij}^6$  の形からも分かるように、 $R_{ij}$  が小さいと  $1/R_{ij}^{12}$  が勝つから、正に発散していくわけである。ということは、ある距離よりも近づけなくなってしまう反発する。 $R_{ij}$  が大きくなると、 $-1/R_{ij}^6$  が勝ってくる。ということは引力が働く。ある距離 (ファンデルワールス半径) よりも近くに入らず、一方遠くは引力が働く、というのが Lennard-Jones 項である。普通、原子間力は大体そういう、ある距離までは引力が働き、近づきすぎると反発するという形をしている。

静電相互作用項というのは  $q_i q_j / R_{ij}$  のことである。ただし、 $R_{ij}$  はふたつの原子間の距離、 $q_i$  は各原子の電荷である。力はポテンシャルエネルギーの位置座標による微分だから、力に直すと、 $q_i q_j / R_{ij}^2$ 。これはクーロンの静電力である。分子としては中性であるが、電子がいろんな所において、大きな原子は小さな原子から電子を吸い取ってしまう。それで水素は、電子を吸い取られて正になる。電子をとってくる、酸素とか窒素等の大きな原子は負になる。各原子が電荷を持っていると考え、それが原子の中心にあるとして、相互作用するという形で近似している。

最後に、ねじれエネルギー項というのは、非常に大雑把な言い方をすると、これらで説明できない力、残りの部分である。本当は量子力学で表せば、これに対応する項を出すことができるのだが、今はすべて古典力学でしている。古典力学で表そうとすると、こういう項は手で入れないといけないわけである。ある特定の角度が

非常に安定していることが、経験的に分かってるわけだが、これを表すことができないから、コサインで表してしまう。

以上が、標準的な蛋白質のエネルギーである。それでこのエネルギーの微分が、それぞれ原子に働く力を与えるわけである。

## § 溶媒の効果

今、蛋白質だけのエネルギーの話をしたのだが、溶媒との相互作用がある。特に溶媒効果が難しい。以下は私の勝手な分類だが、レベルが上がるとだんだん厳密になっていくという意味である。

レベル 1：距離に依存した誘電率  $\epsilon(r)$

(大雑把だが、計算時間は気相の場合とほぼ同じ)

レベル 2：溶媒接触表面積に比例する項の和

(精度は大分まし。計算時間は気相の場合より長い、数倍程度)

レベル 3：溶媒分子をあらわに取り入れる

(厳密だが、計算時間が非常に長い)

現在は私が考える一番厳密な、レベル 3 でやっている。これは水の粒子をあらわに取り入れている。どういうことかということ、水分子  $\text{H}_2\text{O}$  だが、ボンド間の角度が大体 105 度とか分かってるわけである。また、この結合の長さが約 1 オングストロームと分かっている。こういう水分子をたくさん蛋白質のまわりにばらまく。それでこの O と H も、蛋白質の中と全く同じように、静電相互作用と、それから Lennard-Jones 相互作用のエネルギーをもつとする。蛋白質分子も水分子も同じ関数にする。水の場合原子が 3 個しかないの、ねじれエネルギー項は無い。蛋白質と水の原子や、水の原子と水の原子の相互作用を入れるのである。実はそういう項が一番多いので、さきほども述べたように、ほとんど水のシミュレーションになってしまう。

それから、もう少しいい加減にしていくと、まず一番いい加減なのは、誘電率だけで表すというのがある。レベル 1 である。これは何かと言うと、先ほどの蛋白質の原子間のクーロン力に  $\epsilon$  と書いてある。これが誘電率である。真空だと、 $\epsilon$  は 1 になる。ところが、蛋白質の原子の間に水が入っていると、水分子が分極して、原子の電荷を遮蔽する効果がある。それだけクーロンの力を弱めるのである。水があると 1/80 に静電効果が弱まる。それが分かってるので、一番単純に水の効果を取り入れるのは、ここにだけ入れるのである。それで二つの蛋白質の原子間の距離が遠いと、水が充分入っているだろう、近いと水は入れないだろうということで、非常に大雑把だが、その距離によって、近くの場合は  $\epsilon$  は 1 くらい、遠くになっていくと  $\epsilon$  が 80 になるような、S 字型の関数で少なくとも水の遮蔽効果、電気力の遮蔽効果を大雑把に取り入れることができる。そうすると、これはほとんど真空中の計算とほとんど変わらないので楽である。ただ、非常にいい加減な関数である。

レベル 2 というのは、溶媒の接触表面積に比例する項の和ということだが、蛋白

質は入り組んでいるため、狭すぎて水が入っていけない部分がある。まず、蛋白質の各原子を丁度ファンデルワールス半径の球だとして、水を例えば球と仮定してぐるぐる転がしたときに、水の球が接することができる表面積を計算する。ここでまた、この分野で結構発達している、表面積を計算するプログラムがあるのである。ガウスボンネの定理を使って厳密に解析で計算する方法もあるし、それから数値的に、球の上に点をばらまいて、その個数を数えて表面積を計算するなど、いろいろな方法が発達している。それを使って、それぞれの形の溶媒が接している表面積を計算して、その接してる部分が何らかの水の効果を受けるであろうということで表面積に比例する項を入れる。その比例定数は、水の溶媒効果の実験で決めるのである。水はいろいろな効果があるから、静電相互作用しかみていないレベル1よりは大幅に精度は上がるのだが、まだかなりいい加減である。

現在の段階ではわりと、レベル3まで使うことが多い。

### § 極小に留まらない方法

シミュレーションがエネルギー極小状態に留まらない強力な手法を用いる。

それで、エネルギー関数はある程度満足できるものだとして、次に考えないといけないのは、シミュレーションが引っ掛かるのを何とか避けるということである。

それで私たちは主に徐冷法 (simulated annealing) と拡張アンサンブル法 (generalized-ensemble algorithm) の二つの手法を適用することを提唱してきた。徐冷法というのは、結晶をつくるプロセスを利用している。これを計算機上で再現して、温度が高いところからシミュレーションを始めて、徐々に温度を下げていくとエネルギー最小構造に近づく。拡張アンサンブル法はエネルギーの値をランダムウォークさせる。エネルギーが上がったり下がったりするのである。エネルギーが上がる蛋白質はとほどけているし、エネルギーが下がると折り畳まれている。それを何回も繰り返すのだ。エネルギーが下がると引っ掛かってしまう。でも上がることによってまたその引っ掛かりからはずれて、それを繰り返すことによってエネルギー極小状態に留まるのを避ける。この拡張アンサンブル法というのは総称なので、いろいろ特別な方法として、マルチカノニカル法、焼き戻し法、レプリカ交換法など、いろいろある。この拡張アンサンブル法の利点は、一回のシミュレーションで最小エネルギー状態だけでなく、幅広い温度領域における熱力学的量を計算できることである。

### § ミクロカノニカル分布

計算機シミュレーションの手法というのは、アンサンブルを再現する手法である。アンサンブルという、統計力学の概念があって、日本語では統計的集団、統計集団と言う。一言でいえば、アンサンブルというのは、ある調べたい系を与えたときその系がとる状態の集合である。

まず、ミクロカノニカルアンサンブルというのがある。これは何かというと、ミ



クロカノニカルアンサンブルというのは孤立系である。ということはこれ以外に世の中存在しない。これがすべてで、これがすべてなので、外との相互作用は何もない。エネルギーは外へは出ていかないし、外からも入って来ない。完全な孤立系である。孤立系だとエネルギーは保存されて、出ていきもしないし、入ってきもしない。トータルのエネルギーはずっと一定である。こういう系をミクロカノニカルアンサンブルと言う。一つの統計集団である。こういう系は我々のこういう環境ではない。我々の環境は外とのやりとりがあるから、完全に遮蔽されているので外にも出ないし、外から入ってこないのが孤立系なのだ。

## § カノニカル分布

我々の空間というのはむしろカノニカルアンサンブルである。これはまわりに熱浴があって、例えば、この部屋全体、その中に我々が調べたい系があると、この系と熱浴のあいだにエネルギーのやりとりがある。どういうやりとりかということ、系全体の温度を一定にしようということである。だから、この系と外の熱浴とは温度が上がれば出ていくし、温度が下がれば外から入ってくる。こういうのをカノニカルアンサンブルと言う。先ほどの孤立系はエネルギー一定だったが、今度は温度一定である。計算機シミュレーションは、こういった、カノニカルアンサンブルなり、ミクロカノニカルアンサンブルを、再現することができるのである。

## § 作り方

その作り方については、例えばカノニカルアンサンブルだと、メトロポリスのモンテカルロ法というのがある。それはマルコフ鎖というプロセスを用意して、状態を次から次へと、コンピュータ上で作っていく。自動的に温度一定の系に収束するようにすることができて、それをメトロポリス法という。

つまり温度一定の状態を次から次へと発生させる計算機シミュレーション手法が存在するのである。どういうことをやるかということ、乱数をとって新しい状態をつくる。座標を全部指定したら、一つの状態が得られるので、例えば一個の座標を選んで、その座標をちょっとだけ変える。それを新しい状態とする。このときのエネルギーを計算する。エネルギーが下がれば自動的に許すのだが、上がる時にある基準で許す場合と許さない場合と考える。許すかどうかの判定にその温度に依存した関数が入ってくるのである。これがメトロポリス法である。そういう方法を使って、いろいろと計算機シミュレーションをするわけである。

温度  $T$  におけるカノニカルアンサンブルだと、実はエネルギー値  $E$  をもつ状態が実現される確率は  $e^{-\beta E}$  という、いわゆるボルツマン因子というものに比例する。ここで、 $\beta = 1/k_B T$  である ( $k_B$  はボルツマン定数)。こういう確率でそれぞれの状態が出てきて、それぞれのエネルギーに何個状態があるかの個数を掛け算すると、エネルギーのカノニカル分布が得られる。大体一般的にベル状の形をしていて、低

温だとデルタ関数的になる。確率の一番高いところの値が大体平均エネルギーで、その幅がその温度における熱的な揺らぎを表している。

簡単な5個のアミノ酸チロシン、グリシン、グリシン、フェニルアラニン、メチオニンからなるペプチドである Met-enkephalin の真空の系 (ベンチマーク) で、あるエネルギー関数を使うと、この系の最小エネルギー構造は  $-12 \text{ kcal/mol}$  という値をとる (図 6)。 $\text{kcal/mol}$  は、化学でよく使うエネルギーの単位である。物理では電子ボルト (eV) という単位を良く使うが、約  $23 \text{ kcal/mol}$  が  $1 \text{ eV}$  である。 $1 \text{ eV}$  というのは、 $1 \text{ V}$  の電圧で電子一個を加速して得られる運動エネルギーであった。

それで、従来のカノニカルアンサンブルを再現するシミュレーション手法によると、温度が  $1000 \text{ K}$  だと平均エネルギーが大体  $15 \text{ kcal/mol}$  くらいで、分布が (図 7) のようになり、 $600 \text{ K}$  だと (図 8)、 $300 \text{ K}$  だと大体  $0 \text{ kcal/mol}$  あたりが平均値となる (図 9)。ちなみに、エネルギーは定数を足してもいいので  $0$  点に意味は無い。それから  $50 \text{ K}$  ではこのシミュレーションだと  $-7 \text{ kcal/mol}$  くらいで、熱的な揺らぎが非常に小さくなる (図 10)。

それで、ここに計算機シミュレーションの問題が分かるのだが、これは実は間違った答えを出しているのである。先ほどのベンチマークの系だと  $-11$  にならないといけないということが分かっているのだが、計算機シミュレーション、普通のやり方ですと、初期構造に依存しているわけである。どこからシミュレーション始めたかによって違ってしまう。もうこれは熱的揺らぎが小さすぎてエネルギー極小状態から脱出できない。それで間違った答えのほうにずっといくのである。これが従来の手法の問題点で、低温のシミュレーションをやろうとすると、引っ掛かってしまって、間違った答えを出してしまう。自然界は充分時間があるので、いずれちゃんとなる。計算機だと足りないので、引っ掛かってしまう。それで私たちがやったのは、まず徐冷法という方法でやったのだが、マルチカノニカル法、焼き戻し法、レプリカ交換法等もある。

それでとにかく蛋白質の分野にいろいろな手法があるというのはそれだけ問題が難しいということである。できたと思ったらやっぱりダメだった、また次の方法を考えたと思ったらダメだ、次から次へとこの方法はダメだといって新しい方法を考え出す。それだけ、なかなか難しいということを言いたいのである。

## § 徐冷法 (Simulated Annealing)

徐冷法というのはどういう方法かを説明する。高温だと励起エネルギーが大きいので、引っ掛からずにいろいろな構造をとることができる。それで、これは結晶をつくるプロセスをシミュレーションで再現しているので、どういうことをしているかという、シミュレーションをやりながら、徐々に少しずつ温度を下げていくのである。そうするとだんだんとエネルギーが下がって行って、最終的に低温の分布を出すことができる。ここで最小エネルギー構造をとる。

これを急に温度を下げると、熱的揺らぎが小さくなって、どこかで引っ掛かって

しまう。だから、ゆっくりゆっくり温度を下げるのである。徐冷法だと、1000 K からシミュレーションをしながら徐々に温度を 50 K まで下げると、だんだんとエネルギーが下がってきて  $-11$  kcal/mol まで来る (図 11)。

図 11 は先ほどの 1000 K のカノニカルシミュレーション、50 K のカノニカルシミュレーションと徐冷法によるシミュレーションを表している。従来の方法だと 50 K で  $-7$  kcal/mol くらいで引っ掛かってしまったが、1000 K から 50 K までゆっくり温度を下げていくと、 $-7$  kcal/mol で引っ掛からずに、ずっと  $-11$  kcal/mol までいける。結晶をつくる時もそうで、ものを解かして急に水の中につけるときれいにできないが、ゆっくりゆっくり時間をかけると、最後にきれいな結晶ができるのである。

### § 拡張アンサンブル法 (Generalized-Ensemble Algorithm)

一方、拡張アンサンブル法は自然界に存在しない人工のアンサンブルに基づく。ここでは、まず、拡張アンサンブル法の一つである、マルチカノニカル法 (multicanonical algorithm) を説明する。温度  $T$  における従来のカノニカルアンサンブルのエネルギー  $E$  の分布は、状態密度  $n(E)$  かけるボルツマン因子  $e^{-\beta E}$  だった。一方、マルチカノニカルアンサンブルというのは、このボルツマン因子の代わりに、マルチカノニカル重み因子  $W_{mu}(E)$  をかける。この重み因子の定義は、エネルギー分布が一樣であるということである ( $P_{mu}(E) = n(E)W_{mu}(E) = \text{constant}$ )。ということは、 $W_{mu}(E)$  は状態密度分の一であることを意味する。すべてのエネルギーが同じ確率で出てくる。それで、エネルギー空間を行ったり来たりできるわけである。先ほどのカノニカルアンサンブルのベル型だと、ベルの一番上の所の確率と下の所の確率が 10 の何乗も確率が違う。そうすると上の所ばかりしか出てこない。これに対し、マルチカノニカルアンサンブルでは、どのエネルギーも同じ確率で出てくる。

ただ、状態密度、つまりあるエネルギーの時に状態が何個あるかという情報は、実は分からないことが多い。イジング模型のように厳密に解けている問題はこれが分かっているということなのだが、普通の自然界の系は大部分はこれが分からない。だから、最初にこの推測値を大雑把に求めようというのである。

まず始めに、短いシミュレーションを、何回も繰り返す。それによって、マルチカノニカル重み因子  $W_{mu}(E)$  の大体の値を求める。ある程度のものが見つかったら、長い Run をやって、その結果を解析することによって、エネルギーがどう変化するかを見ていくと、最小エネルギー構造が分かるし、それからどういうふうに変化したか状態をためておいてそれをあとで解析すると、実は任意の温度のカノニカル分布が再現できる。これが、今回のシミュレーションで得られているエネルギー分布を、これを発生させた重み  $W_{mu}(E)$  で割って、それでさらに普通のボルツマン因子をかければ、任意の温度  $T$  における自然界の情報が再現できる。

つまり、マルチカノニカルというのは、あの世に行って、自然界に存在しないも

のを自分で作ってしまっているわけである。あっちでやっておいて、最後にこの世の情報を引き出す。ちょっと比喩が良くないかもしれないが、完全に人工の、自然界に存在しないアンサンブルをつくってしまってるわけなのである。

まず、普通の高い温度のカノニカルシミュレーションを一回やって、その分布、つまりそれぞれのエネルギーの値の時、何個構造が存在したか、個数を数えるわけである。それをヒストグラムで数えて、その対数を重み因子に足して、新しい重みをつくる。この重み因子というのが先ほどのシミュレーションのエネルギーが上がったときにアクセプトするか、捨ててしまうかの判定の関数である。実はボルツマン因子にしたがって、エネルギーが上がったときに、採用するか捨ててしまうか判定しているわけだが、その判定する関数を変えていくわけである。1回目のシミュレーションは、カノニカルシミュレーションをだすボルツマン因子の判定条件でやって、2回目はちょっとそれを変えてやると、ちょっとフラットな分布が得られる。これを繰り返すと、5回目でかなりの部分、フラットになるのである。この Met-enkephalin のような簡単な系だと、大体5回ぐらいの短いシミュレーションでフラットにできる。そうするとその間を行ったり来たりするわけである。

それでどういうことが起こるかということ、エネルギーが上がったり下がったりということが起きる。 $-12 \text{ kcal/mol}$  が最小エネルギー値だったが、最小エネルギーにも到達するし、 $1000 \text{ K}$  の平均エネルギーの  $15 \text{ kcal/mol}$  も越えるし、非常に幅広くカバーするシミュレーションが実現する。最小エネルギーの値の時にどんな構造だったか保存しているので、このときの最小エネルギーの構造を求めたら、一番エネルギーの低い立体構造が得られる。

この手法には大きなメリットがある。何かというと、徐冷法だと途中まで下がってくると、もう引っ掛かりやすくなってしまふ。そして、もしあるところで引っ掛かってしまうと、ずっと引っ掛かっている。マルチカノニカル法は、引っ掛かってもまた上に行って、それを何回も繰り返すので、この徐冷を何回も繰り返す様な事に対応している。それよりもさらに良いことは、徐冷法というのは常に平衡を破っている。温度をどんどん変えていくから、常に平衡を破っている。それでこのシミュレーションというのは長いシミュレーションをやって十分な平衡の情報を使って初めてその温度の情報が得られるのだが、常に平衡を破っているので、例えば  $1000 \text{ K}$  から  $50 \text{ K}$  まで温度を下げたとき、 $1000 \text{ K}$  から  $50 \text{ K}$  の間の温度における平均値とかは案外いい加減なものになってしまう。ところがマルチカノニカルシミュレーションは常に平衡を保ったままやっている。そういうことで平均値を出すとき、高温も途中の温度の熱力学量も非常に精度良く出すことができるのだ。そういう意味で、最小値を求めるのは徐冷法でも充分かも知れないけれど、全体像を知ろうとするのには、こちらの方がはるかに信頼性が高い結果を出すことができる。

拡張アンサンブルというのはエネルギーとは限らず、ある変数を選んでその軸に沿って行ったり来たりできるように、人工的にやるという事である。そうすると、行ったり来たりすることによって引っ掛かりを避けることができる。引っ掛かって間違った答を出すのを避けて、正しい系の全体像を見ることができるのだ。

## § レプリカ交換法 (Replica-Exchange Method)

拡張アンサンブル法のもう一つの例であるレプリカ交換法を説明しよう.

### 1. 系

互いに相互作用しない異なる温度を持つオリジナルシステムのコピー (レプリカ) を  $M$  個用意.

### 2. レプリカ交換

各レプリカで独立にカノニカルシミュレーションを実行.  
途中で温度値が隣接した二つのレプリカの温度を交換.  
この二つの操作を繰り返す.

### 3. 任意の温度のカノニカル分布

多ヒストグラム再重法を用いて計算.

もう一つはレプリカ交換法というのがあるのだが, これは, 今まで一個の系でやっていたのだが, わざと  $M$  個コピーを用意する. 相互作用しない, 完全に独立な系を用意するのである. 何をやるかという, 違う温度を用意しておく.  $M$  個違う温度で平行して走らせる. その途中で温度を入れ替える. これはもう完全に自然界では有り得ないことだが, 勝手に独立に従来のカノニカルシミュレーションを違う温度で走らせて, 途中で温度をどんどん入れ替える.

そうすると, 一つの系を見ていると, 温度がぐちゃぐちゃに動き回る. 温度が上がるとエネルギーが上がるし, 温度が下がるとエネルギーが下がるので, エネルギー状態のランダムウォークを引き起こして, エネルギー極小状態に引っ掛かるのを避ける.

それぞれ違う温度で走らせていて, あるところで温度を引っくり返す. これも実はモンテカルロなのである. メトロポリスの判定でやっている. アクセプトされれば, レプリカの温度が入れ替わり, リジェクトされるとそのまま同じ温度で行く. これを何回もやっていると, 一つのレプリカを見たら温度がいろんな値をとっている. 温度が上がったり下がったりして, 上がると蛋白質がほどけて, 下がると折り畳まれて, これを何回も繰り返す. それで例えば一つの温度をずっと追っていくと, それはその温度のカノニカル分布が得られる. レプリカを一つ決めて追っていくと温度が変わっていくが, 温度を一つ決めて追っていくと, その温度のカノニカル分布が得られるのだ.

それで, ある一つのレプリカを見ると, 温度が上がったり下がったりシミュレーションの途中であるわけである. この温度の上がり下がりにしたがってエネルギーも上がったり下がったりする.

それでレプリカ交換法では, 充分な分布の重なりがあると交換が起こる. 判定条

件はこのエネルギー差で決まるから、この方法は重みを決める必要がない。先ほどのマルチカノニカル法というのは、マルチカノニカル重み因子を最初に決めないといけないのだが、それが複雑な分子だと非常に難しくなる。このレプリカ交換法というのは、普通の一定温度のシミュレーションをやればいいだけである。ところが、こっちの方は系のコピーがたくさんある。マルチカノニカルシミュレーションは一個の系ですが、これは  $M$  個の系、 $M$  個のコピーを用意するので  $M$  倍の計算時間が必要である。例えば、アミノ酸が数十個の小蛋白質の系でも、100 レプリカぐらい必要になる。その結果、100 倍の計算時間が必要になる。重み因子を決める必要は無いが、個数がある。そういうふういろいろな方法に一長一短があるわけである。

### § 蛋白質の折り畳みシミュレーションの例

先ほどのベンチマークの Met-enkephalin の系だが、ランダムウォークした結果を重みを決め直して正しい自然界の分布を出すと、一回のシミュレーションから温度の関数として平均エネルギーを出すことができるわけである。先ほど 50 K で  $-7$  kcal/mol に引っ掛かっていたが、50 K だと実は  $-11$  kcal/mol が正しい値である。従来のやり方だと引っ掛かってしまうけれど、マルチカノニカルシミュレーションだと一回のシミュレーションで幅広い範囲の温度の平均値を出することができる。従来のやりかただと、一つの温度でシミュレーションをやって平均値を出すわけであった。次の温度でまた平均値を出す。それを一回で全部出してしまふ。

次の例は、アミノ酸のホモポリマーである。アラニンばかりのホモポリマーやグリシンばかりのホモポリマーなどについて、一回のシミュレーションの結果から、何パーセントの割合でヘリックスができるかなどが分かる。ホモアラニンでは、低温で 80 パーセント、ヘリックスができるが、高温だとヘリックスはほどける。このようにアラニンというアミノ酸ばかり集めたものはヘリックスを非常に作りやすいのだが、グリシンというのは今度はヘリックスをつくらない。こういう情報を一回のシミュレーションで出すことができるのである。

### § 水を入れたシミュレーション

それで、エネルギー関数の時に言ったが、レベル 1、レベル 2、レベル 3 と水の効果の精度をだんだん上げていくとどうなるかという話だが、今までの結果は全部真空中の結果である。水が無いから簡単である。簡単でも引っ掛かるのだが。水を入れるとどうなるかという、様変わりしてしまう。図 12 はエックス線の実験で得られた、C ペプチドという 13 個のアミノ酸の系の構造だが、水無しでこの系をシミュレーションすると最小エネルギー構造が、ヘリックスの部分は出てくるのだが、全体的には大分違った形になっている (図 13)。レベル 1 の水を入れる。レベル 1 は大分大雑把だが、真空よりは似てくる (図 14)。さらにレベル 2 の表面積に比例する項

を入れると、図 15 のようになる。こういうふうに水の寄与が重要だという事が分かる。

もう一つの例 (図 16) だと、ベータシートをつくっているが、真空中でやると、図 17 が最小エネルギー構造だが、レベル 1 ではヘリックスができてしまう (図 18)。これはまずいという事でレベル 2 ですとベータシートが得られる (図 19)。

いかに溶媒の寄与の近似が少し変わると結果が変わってしまうかということが分かる。

先ほどの五個のアミノ酸の Met-enkephalin では、図 20 が実は真空中の最小エネルギー構造である。こういう構造をベータヘアピンという。図 21 は次に安定な構造である。水素結合が 2 番目と 5 番目だったのが、1 番目と 4 番目になっている。こういうふうに真空中だとベータヘアピン構造をしているが、水を入れるとどうなるかと言うと、延びてしまう。真空中だとヘアピンだったのが水が間に入り込んで伸ばしてしまうのである。

図 22 は末端間距離と言って、蛋白質の端から端までの直線距離である。これが小さいとコンパクトになっているし、延びているとほどけているという事である。そうすると真空中でやると 200 K では末端間距離の平均値が 5 オングストロームぐらいである。ところが温度が上がってくるとほどけてくるから、平均の末端間距離が延びてくる。ところが水中だと全ての温度ですっと延びたままである。これは先ほど言った、水が間に入りこんで伸ばしてしまうということである。いかに水が大事か、水によって全然違うということをいっている。

先ほどレベル 2 までやってヘリックスのそっくりな構造になった、C ペプチドという 13 個のアミノ酸の系をレベル 3 でやってみる (図 23)。水を 1387 個球の中に入れて、その中に C ペプチドを一個入れる。トータルの原子数は、4365 個である。この系は実験で 50 パーセントから 60 パーセントヘリックスをつくることが知られている。従来のカノニカルシミュレーションでやるとエネルギーはあまり変わらず、引っ掛かっているところがなかなか動かない。我々の方法だと 5000 kcal/mol の間をランダムウォークするのである (図 24)。

## §RMSD と慣性半径

実際これで構造もかなり変わっている。エネルギーだけ変わってるのではないかという心配をする人があるが、構造も変わっている。なぜエネルギーだけ変わっているのを心配するかというと、水がこんなにあると、数個の水がちょっと動いただけでエネルギーがはね上がる。ちょっと近づけば、反発するので、エネルギー値としてはちょっと水が動くだけで変わってしまうのである。それなら、構造自体はあまり動いてないんじゃないかという心配があるのだが、実際にエネルギーのランダムウォークとともに動くことが分かる (図 25)。

RMSD というのは root-mean-square deviation と言って、エックス線の実験で得られた構造からどれくらい座標がずれたかを表す。つまり、二つの構造を比べると

きに、まず平行移動して重心を重ねる。あと回転してできるだけ二つの構造を重ね合わせる。そうするとどうしても合わない、対応する原子間距離の自乗の平均とったものがRMSDである。だから完全に平行移動して重心を重ねて、回転して、完全に二つの構造が一致したら0になる。0に近いほど構造が近いわけである。先ほどのエネルギーがランダムウォークしてる間にどれくらい自然の構造とそっくりかというのをみると、全然違っていたのである。

もう一つは慣性半径と言って、蛋白質全体を一つの球と見たときの球体の球の半径である。ほどけているとそれは大きいし折り畳まれていると小さい。それも先ほどのエネルギーがランダムウォークしている間に、小さくなったり大きくなったりしている。この二つのものを見れば構造が大分変わっていることが分かる。それを視覚的に見たのが先ほどのやつである。エネルギーが変わるだけでなく形も変わっている。そういう意味で非常に構造サンプリングがうまくいっている。

## § 自由エネルギーの地形

それで蛋白質の折り畳みに結局自由エネルギーの地形を見るということが重要になってくる。先ほどエネルギーと言っていたけれども、実は自由エネルギーを見ないといけない。エネルギーの最小値と自由エネルギーの最小値は違っていて、本当は自由エネルギーの方で最小値を求めないといけないのである。エネルギー  $E$  と、自由エネルギー  $F$  の間には、

$$F = E - TS$$

という関係がある。ここで、 $S$  はエントロピーである。実はエントロピー効果もあるわけである。自由エネルギーで本当は見なくてはいけない。それがどういう形をしているかという話だが、蛋白質の系が折り畳むことができるというのは、自由エネルギーの形が図26のような漏斗状をしていないといけない。こういう傾きを持っているから転がり落ちて一つの構造に折り畳まれるのである。

それを二つのパラメータで記述することができる。何かというと、ガラス転移温度 ( $T_G$ ) と折り畳み温度 ( $T_F$ ) である。結局、蛋白質の折り畳みが二つのパラメータで表されてしまうのである (図27)。折り畳み温度というのは、自然の構造に折り畳まれる温度である。蛋白質でどういうことが起こるかということ、まず温度が高いとランダムコイルでぐちゃぐちゃしていて、温度を下げてくるとこれがコンパクトになる。こういうランダムコイルから球状になる転移温度をコイルグロビュール転移温度と言う。だから最初温度が高いときは延びていて、温度を下げるとコンパクトになるけど、コンパクトの中で動き回っている。さらに温度を下げると自然の構造になる。自然の構造が得られる温度を折り畳み温度というのである。ガラス転移温度というのは何かというとどこかに引っ掛かってしまう温度である。温度が低くなってくると、熱的揺らぎが小さくなって、エネルギー極小状態に引っ掛かってしまう。結局蛋白質が折り畳むことができるっていうのは、ガラス転移温度が折り畳み温度よりも小さくないと、もし高いところで引っ掛かり出したら、一つの所に行



くことができないのである。

他の言い方で言うと、図 27 に  $\delta E$  と  $\Delta E$  というのがあるが、 $\delta E$  というのは、一番自由エネルギーが低いところと他のところの構造とのエネルギー差である。だから自然の構造へ折り畳んでいこうというエネルギーギャップだから、自然の構造に落ち込ませようという力ということである。 $\Delta E$  というのはそれぞれのエネルギー曲面の凸凹を表している。これは引っ掛かりやすさである。引っ掛かりやすさ、即ち凸凹より、一つの構造に落ち込もうとする力の方が大きければ、折り畳まれる。凸凹が大きかったら、いくら行こうとしても引っ掛かってしまう。このように蛋白質の系を二つのパラメータで表すことができる。折り畳み温度とガラス転移温度、またはこういうエネルギーギャップと凸凹の引っ掛かりやすさである。

真空だと、5 個のアミノ酸の場合はコイルグロビュール転移温度が大体 300 K で、折り畳み温度が 230 K である。我々の手法だと、一回のシミュレーションで出すことができる。それからガラス転移温度は 150 K よりも下。ガラス転移温度が折り畳み温度よりも充分下なので、5 個のアミノ酸は折り畳まれているのである。

全体的に自由エネルギーの地形を書くことによって、それぞれの蛋白質の状態を調べることができる。

## § エネルギーの比較

それで、いままで蛋白質のエネルギー関数は、いま二種類くらい使って話してきたのだが、実はこれくらい種類がある。

AMBER というのに 94 年のバージョン、96 年のバージョン、99 年のバージョンがあり、CHARMM, OPLS, GROMOS といった具合である。これくらいエネルギーのパラメータが違うのである。AMBER はカリフォルニア大学サンフランシスコ校の人たちが開発した。CHARMM はハーバード大学の人たち、OPLS はイェール大学の人たち、GROMOS はオランダの Groningen 大学の人たちが開発したパラメータである。ここでパラメータといっているのは、エネルギー関数の定数(数値)である。

こういうパラメータが、グループによっていろいろ違って、実は五つくらい関数、パラメータがある。それがいいか悪いかの判定は誰もちゃんとやったことがなかった。それぞれが自分達がつくったものが正しいと思っていたのである。実は、比べようがなかったのだ。何故かという、彼らはパラメータをアミノ酸一個ごとに決める。アミノ酸をつないだときに本当にそれでいいのかどうか、分からないわけである。比べようと思ったらシミュレーションをやらないといけないのだが、従来のやり方だったら引っ掛かってしまうので、例えばエックス線で決められた構造からスタートすると、そこから動かないのである。それでどのパラメータを使っても動かないから、どちらが正しいかなかなか言えないのである。でも幸い、我々は拡張アンサンブル法でぐちゃぐちゃにかきまわすことができたので、自然の構造からスタートするのではなくて、ぐちゃぐちゃの所からスタートして、折り畳むことがで

きるかどうかの判定ができる。それによって初めて、これらのパラメータが、世界に存在する五つのパラメータが、どれがいい悪いという判定ができるようになったのだ。

我々の基準は次のようなものである。Cペプチドは、実験ではヘリックスをつくることが知られている。一方、Gペプチドは、実験的にベータシートをつくることが知られている。だからぐちゃぐちゃのランダムな初期構造からシミュレーションを始めて、Cペプチドでヘリックスができて、Gペプチドでベータシートができるかどうか。それだけである。この単純なクライテリオンで、これらを比べてみる。そうすると面白いことに全然違った結果が出てくるのである。

例えばGROMOSだと、Cペプチドでもヘリックスがほとんどできない。むしろベータヘアピンができる。

Gペプチドでは、ベータヘアピンができないといけないのだが、AMBER94ではすごいヘリックスができてしまう。これはダメだと、こういう感じでどんどん殺していくわけである。そういうことをやった結果は（未だに途中結果であり断定できないが）図28のようになる。

## § 膜蛋白質

今までは水溶性の蛋白質の話をしてきた。実はゲノムを見ていると、水溶性の（水に溶ける）蛋白質の個数は全体の約3/4で、残りの約1/4は膜蛋白質なのである。ところが膜蛋白質というのは実験で構造を決めるのが非常に難しい。現在実験で決められた蛋白質の立体構造のデータベース（Protein Data Bank）には、約二万個の構造が保管されているが、膜蛋白質は、この二万個の中の数十個しか決まっていなのである。残りは全部水溶性である。水に溶ける蛋白質なのである。

膜蛋白質というのは膜を通り抜けている蛋白質で、大体アルファヘリックスになっている。それでいろいろなデータベースを使うプログラムがあって、ゲノムのアミノ酸配列の情報から、これが水溶性蛋白質なのか膜蛋白質なのかを判定をすることができる。これが99パーセントの確率で当たる。それで膜蛋白質だと判定されたら、今度はその配列の中でどこが膜貫通のアルファヘリックスの部分かというのを推測することができる。それが96パーセントの確率で当たる。だからそれを信用して使うことにする。

今度は大分いい加減な、ヘリックスの部分は一本の筒で剛体として扱い、平行移動と回転を考える、また、側鎖の部分は普通に動かすという形でシミュレーションする。レプリカ交換法で、エネルギーの小さいところを探すわけである。これで得られた最小エネルギー構造を取り出すと実験の構造とそっくりになる。

誘電率が何になるかは重要で、膜蛋白質は膜があり、膜の部分の誘電率は4ぐらいだと言われているので、4だと思っていたのだが、よくよく考えると、ヘリックスとヘリックスの間は狭くて、膜の脂質はあまり入り込めないのである。ということはやはり誘電率として真空の値を使うべきだと考えられる。そうすると実験で決

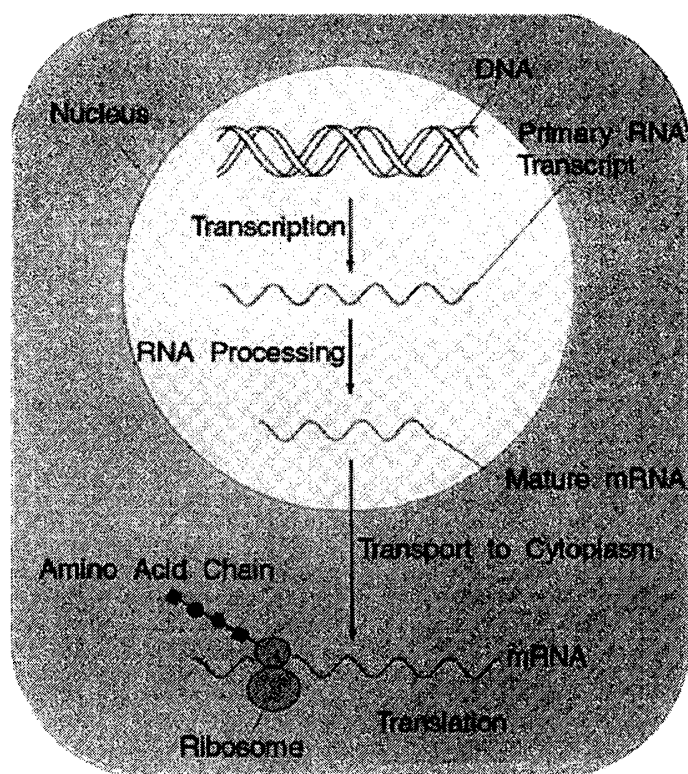
められた構造とそっくりになるのである。一応それでいいんだと理解が得られた。

現段階で実験で決められた構造は数十個しか無いのだが、それをどんどんこのやり方で、理論的に求めてしまおうということもしている。

最後に、本講義の内容の詳細については、例えば、次の文献を参照されたい。  
岡崎進・岡本祐幸（編）化学フロンティア No.8 「生体系のコンピュータ・シミュレーション」（化学同人, 2002）。

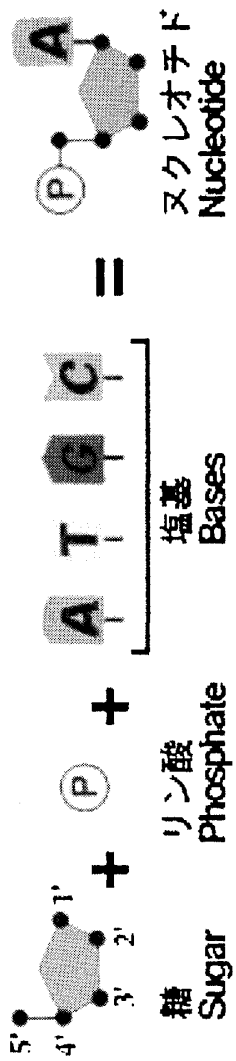
図 1

## 細胞



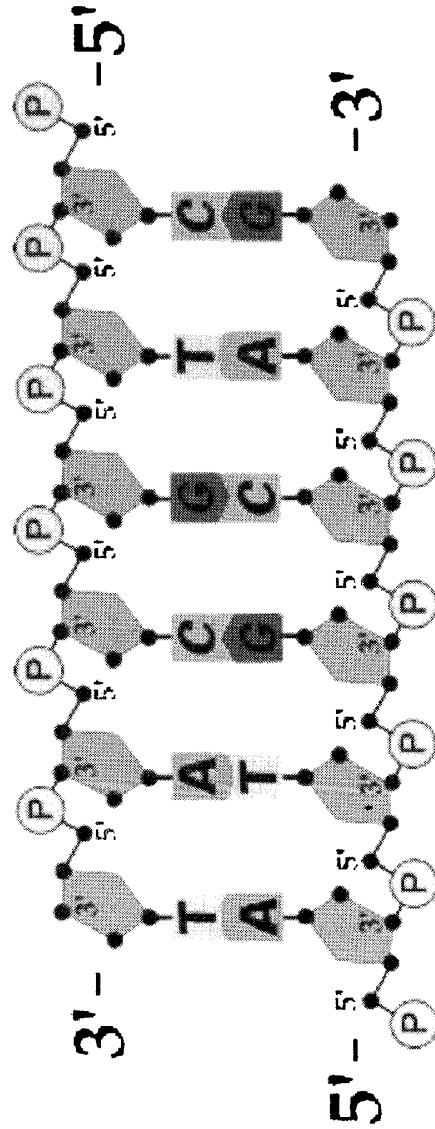
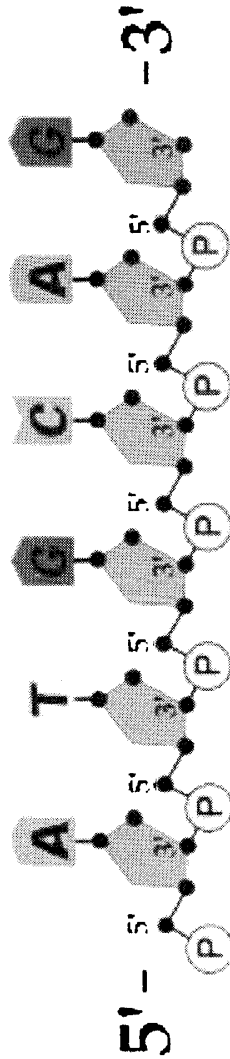
遺伝情報は細胞の核内に存在する染色体に書き込まれている。そして、染色体はDNAの2重らせんによって構成されている。

図 2



# DNA

A: アデニン  
 G: グアニン  
 C: シトシン  
 T: チミン  
 (RNAでは  
 U: ウラシル)



## 2重鎖

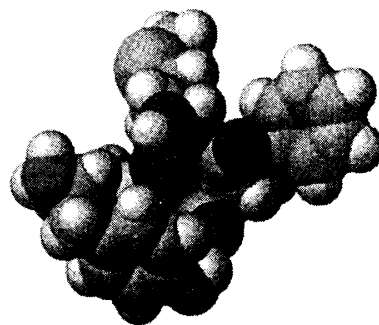
図 5  
IMMUNOGLOBULIN  
(N=216)



199

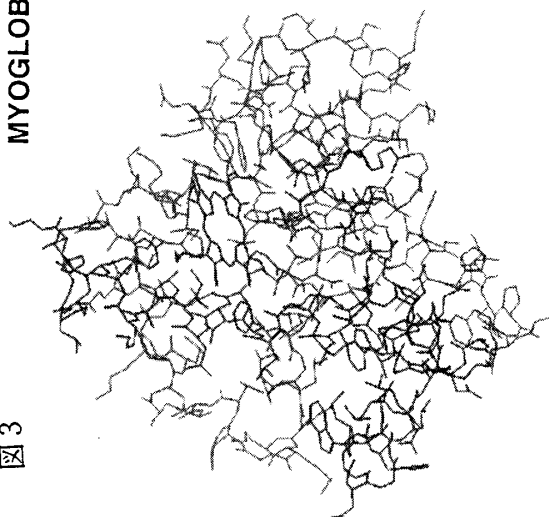
図 6  
Met-Enkephalin:  
真空中の最小エネルギー構造

Tyr-Gly-Gly-Phe-Met  
(N = 5)



E = -12 kcal/mol

図 3  
MYOGLOBIN (N=153)



Mb-1

図 4  
MYOGLOBIN (N=153)



$\alpha$ -HELIX

Mb-2

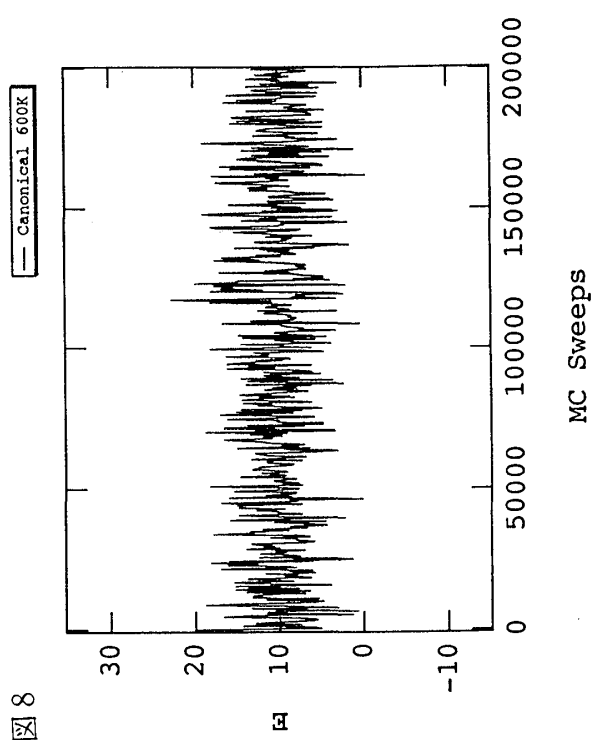
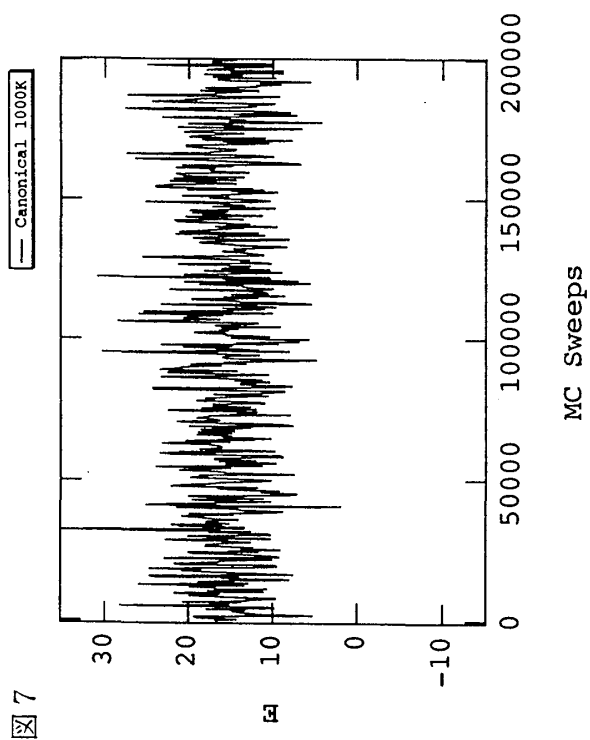
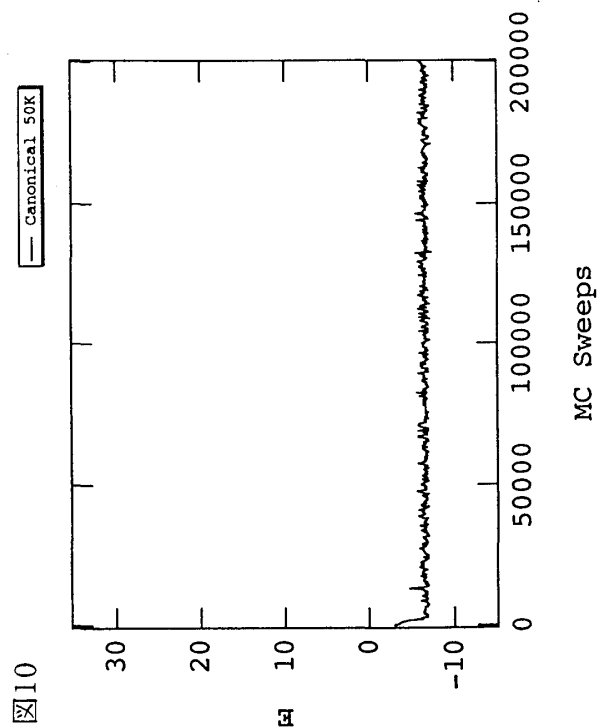
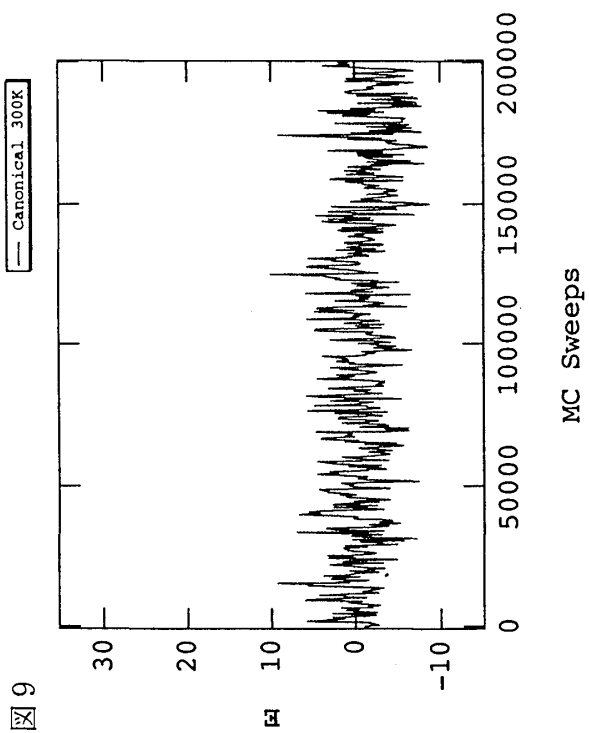
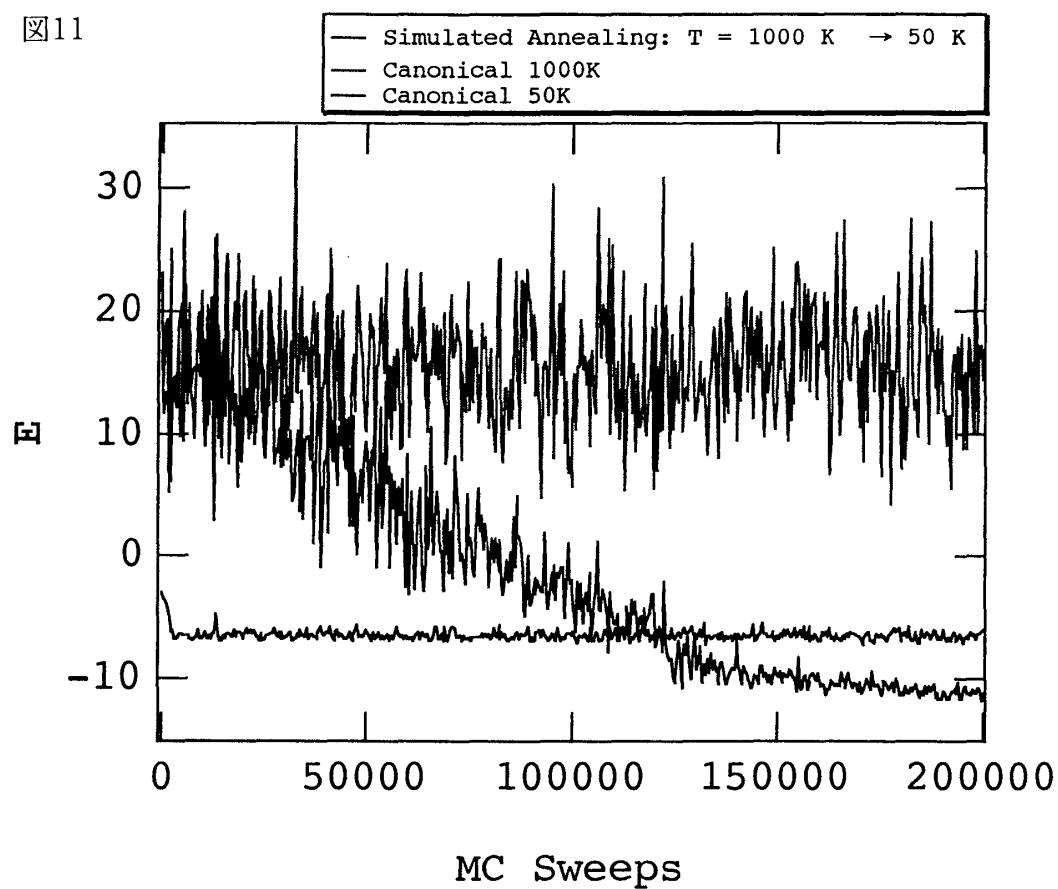


図11



C-peptide-4

図12 C-Peptide (N = 13): X-Ray



図13 Level 0 (Gas Phase)

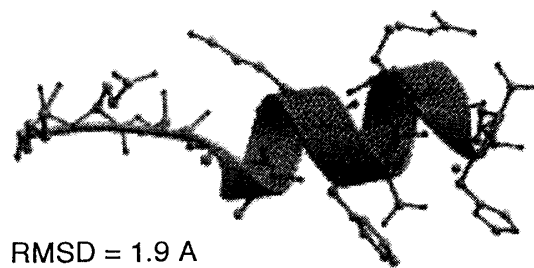


図14

Level 1

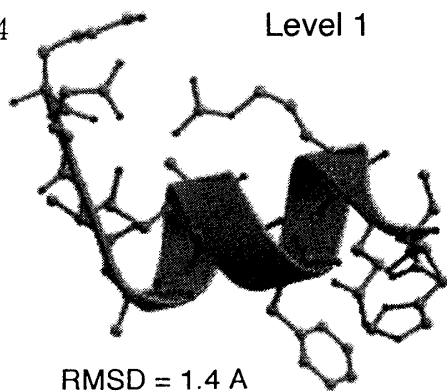
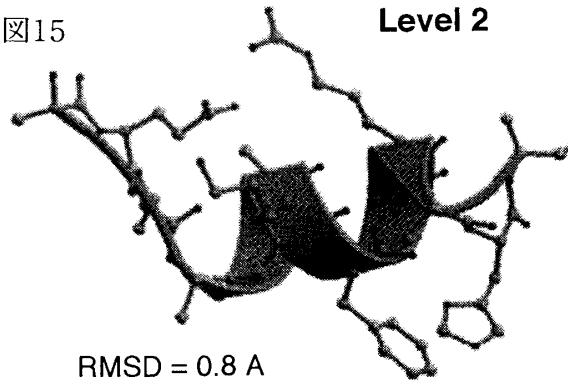


図15

Level 2



M. Masuya, T. Nakazawa & Y.O., in preparation.

BPTI-4

図16 BPTI(16-36) (N = 21): X-Ray

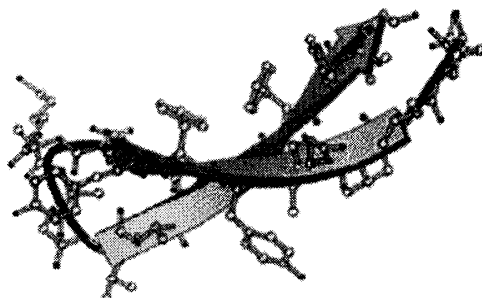


図17 Level 0 (Gas Phase)

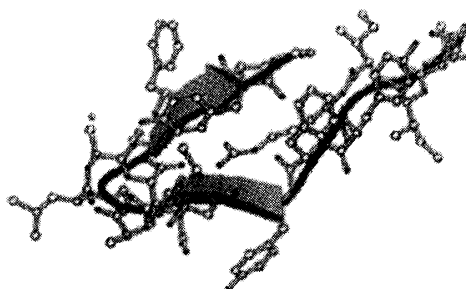
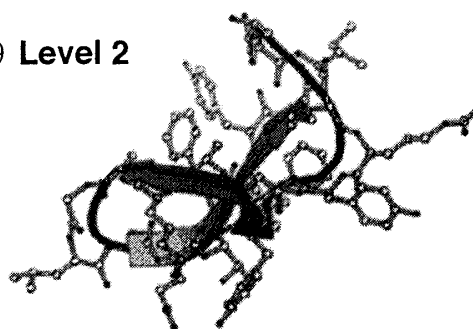


図18 Level 1



図19 Level 2



Y.O., M. Masuya, M. Nabeshima & T. Nakazawa, *Chem. Phys. Lett.* **299**, 17 (1999).

## Two Lowest-Energy Structures of Met-enkephalin in Gas Phase

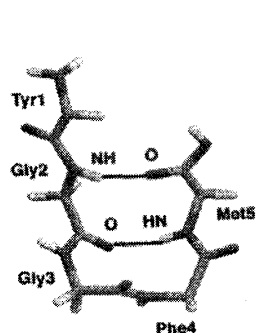


図20 Structure A (Lowest Energy)

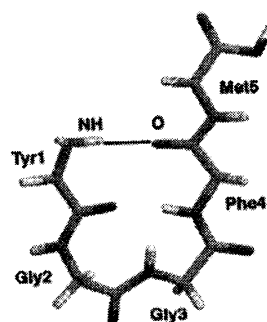


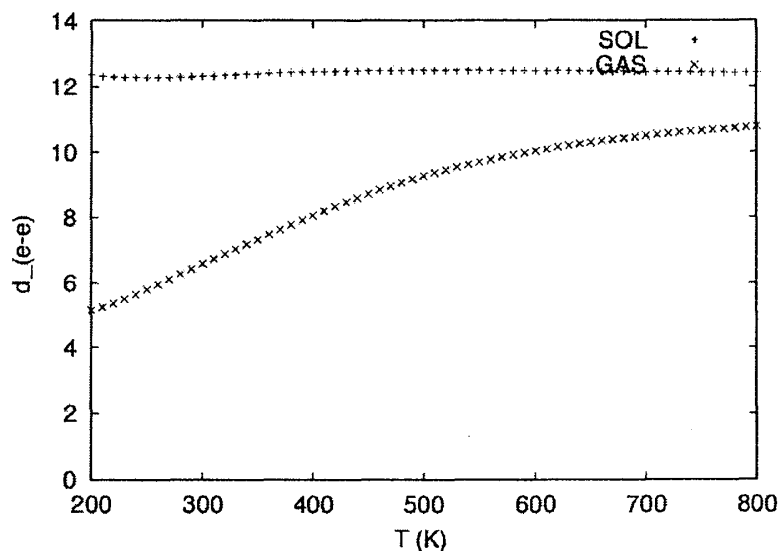
図21 Structure B (Second Lowest Energy)

See A. Mitsutake, U. Hansmann & Y.O., *J. Mol. Graphics Mod.* **16**, 226 (1998), for details.



## Average End-to-End Distance of Met-enkephain in Gas Phase and in Aqueous Solution (RISM): ECEPP/2

図22



A. Mitsutake, M. Kinoshita, Y. O. & F. Hirata,  
*Chem. Phys. Lett.* **329**, 295 (2000).

図23

## C-Peptide of Ribonuclease A (N=13) in Water (Level 3)

Amino-Acid Sequence:

ACE-Ala-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Leu-Arg-Ala-His-Ala-NME

AMBER 99 Force Field (peptide)  
TIP3P (water)

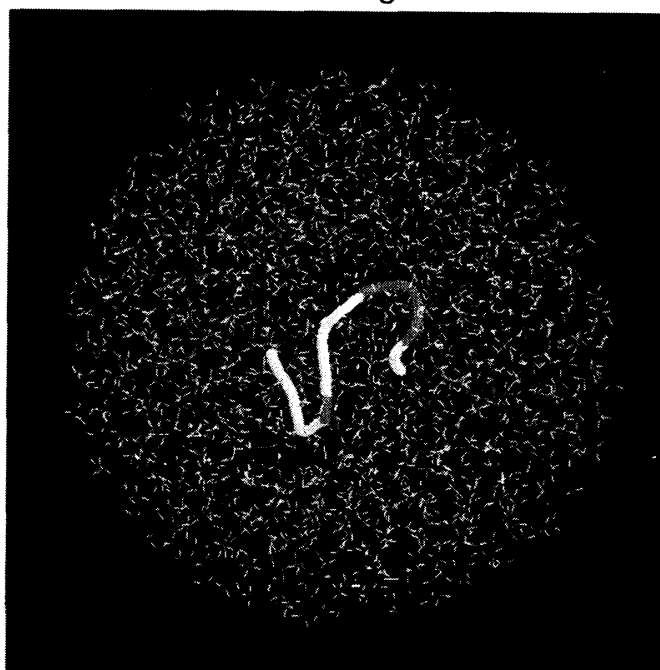
13 residue peptide

1387 Water Molecules

(R = 22 Å)

Total 4365 atoms

Initial Configuration



### NMR Studies (Osterhout 1988)

50-60% helix in solution

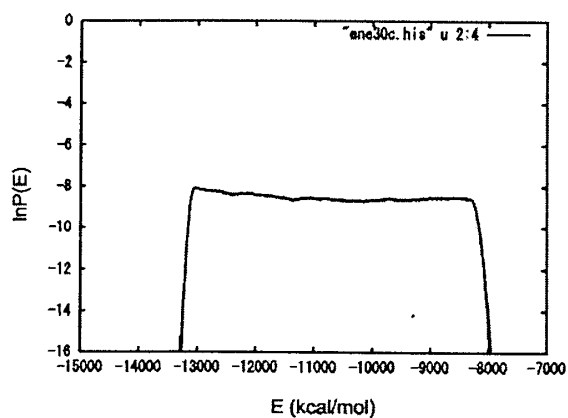
First three residues are  
distorted

Formation of a salt bridge  
between Glu2 and Arg10

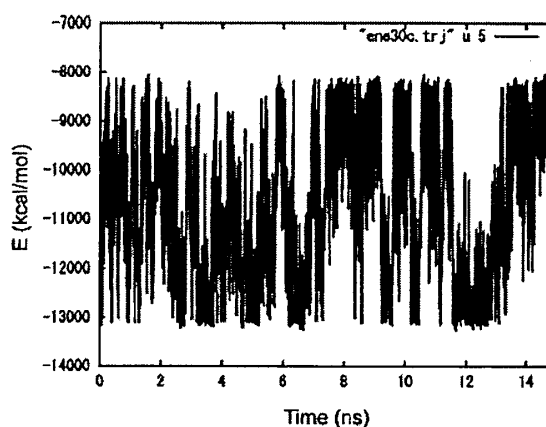
図24

# C-Peptide of Ribonuclease A (N=13) in Water

3. REMUCA (1 replica)  
15 ns



Time series of  
potential energy



# of tunneling events = 46  
(-12850, -8250)

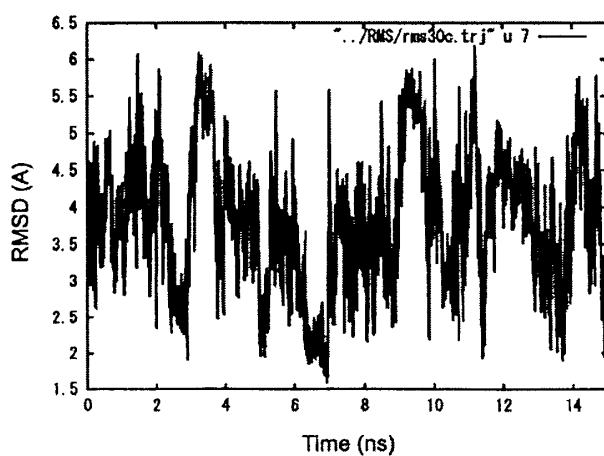
図25

# C-Peptide of Ribonuclease A (N=13) in Water

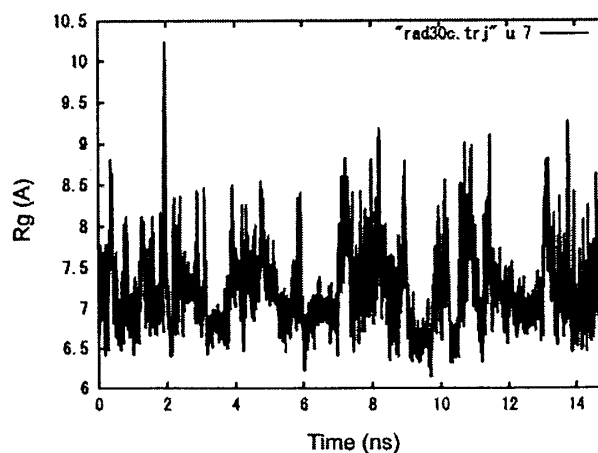
RMSD from "native structure"

Rg: radius of gyration

Time Series of RMSD



Time Series of Rg



## 図26 タンパク質の系の自由エネルギー地形

ガラス的 (Glass-like)

漏斗状 (Funnel-like)

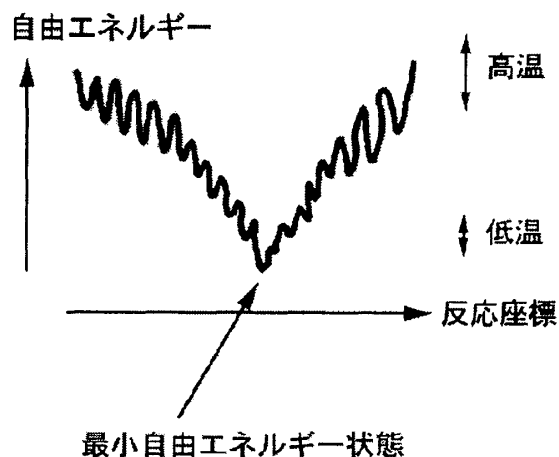
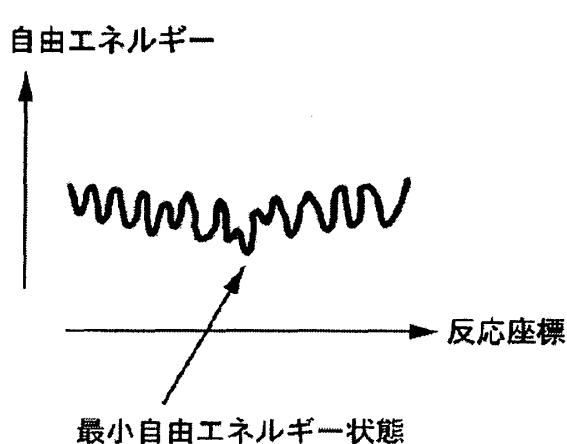


図27

## Folding Funnel and Free Energy Landscape

Cf. J. Bryngelson & P. Wolynes, *Proc. Natl. Acad. Sci. USA* **84** (1987) 7524.

J. Bryngelson, J.N. Onuchic, N.d. Socci & P. Wolynes, *Proteins* **21** (1995) 167.

Principle of Minimal Frustration:

$$T_G < T_F$$

Or,

$$\Delta E < \delta E$$

Cf. Consistency Principle:

N. Go, *Ann. Rev. Biophys. Bioeng.* **12** (1983) 183.

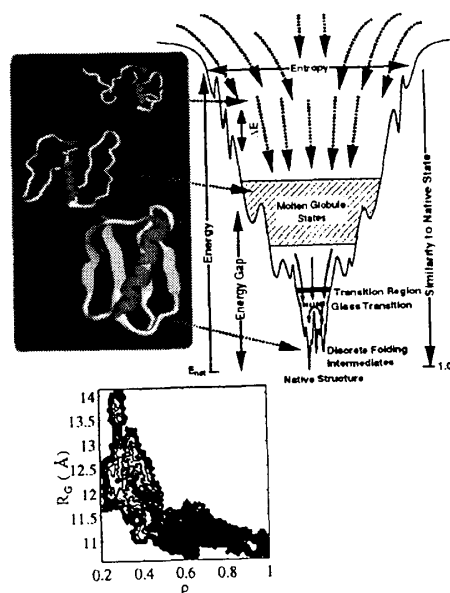


図28

T. Yoda, Y. Sugita, &amp; Y.O. in preparation.

## Comparisons of Various Force Field Parameters (Preliminary Results)

	C-Pept ( $\alpha$ -helix)	G-Pept ( $\beta$ -hairpin)	判定
AMBER			
parm94	○ ( $\alpha$ -helix)	× ( $\alpha$ -helix)	×
parm96	○ ( $\alpha$ -helix)	○ ( $\beta$ -hairpin)	○
parm99	○ ( $\alpha$ -helix)	× ( $\alpha > \beta$ )	×
CHARMM	○ ( $\alpha$ -helix)	○ ( $\beta > \alpha$ )	○
OPLS-AA/L	○ ( $\alpha$ -helix)	○ ( $\beta$ -hairpin)	○
GROMOS	× ( $\beta$ -hairpin)	○ ( $\beta$ -hairpin)	×